The Outline of an 'Intelligent' Image Retrieval Engine

Mohammed Belkhatir, Philippe Mulhem, Yves Chiaramella

THE OUTLINE OF AN 'INTELLIGENT' IMAGE RETRIEVAL ENGINE

Belkhatir Mohammed Mulhem Philippe Chiaramella Yves
CLIPS-MRIM/IMAG
Rue de la Bibliothèque, St Martin D'Hères, France
{belkhatm,mulhem,chiara}@imag.fr

ABSTRACT

The first image retrieval systems hold the advantage of being fully automatic, and thus scalable to large collections of images but are restricted to the representation of low-level aspects (e.g. colors, textures…) without considering the semantic content of images. This obviously compromises interaction, making it difficult for a user to query with precision. The growing need for 'intelligent' systems, i.e. being capable of bridging this *semantic gap*, leads to new architectures combining multiple characterizations of the image content. This paper presents SIR[1], a promising high-level framework featuring semantics, signal color and spatial characterizations. It features a fully-textual query module based on a language manipulating both boolean and quantification operators, therefore making it possible for a user to request elaborate image scenes such as a *"covered(mostly grey) sky"* or *"people in front of a building"*.

KEYWORDS

Multimedia, Image Retrieval, Automatic Extraction, Conceptual Graphs, Image Query Language.

1. INTRODUCTION

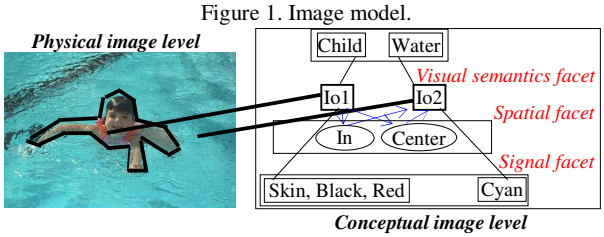
The democratization of digital image technology has led to the need to deal with a new generation of image retrieval frameworks combining expressivity, increased retrieval performance and computational efficiency.
The first content-based image retrieval (CBIR) systems (signal-based) [12] propose a set of image indexing methods based on low-level features such as colors, textures, geometrical forms... which are fully automatic and able to process queries quickly. However, the problem arising from low-level characterization lies on the loss of semantic information conveyed by the image. For example, can we accept that our system considers red apples or Ferraris as being the same entities simply because they present similar color histograms? Definitely not, as shown in [8], taking into account aspects related to the image content is of prime importance for efficient photograph retrieval.
State-of-the-art systems which attempt to deal with the semantics/signal integration [6,9,14] are based on the association of textual annotations with relevance feedback. Prototypes such as iFind [9] and ImageRover [6] offer loosely-coupled solutions based on textual annotations to characterize semantics and on a relevance feedback scheme operating on low-level signal features. These approaches present two major drawbacks: first, they lack to exhibit a single framework unifying signal features and semantics, which penalizes the performance of the system in terms of retrieval efficiency and quality. Then, the user is to query both textually in order to express high-level concepts and through several and time-consuming relevance feedback loops to complement his initial query; which does not enforce a user-friendly interaction.
We propose a multi-facetted image indexing and retrieval framework unifying semantics, signal color and spatial characterizations. We first specify an automated framework extracting the visual semantics. We then enrich the description of images through the specification of processes establishing a correspondence between extracted low-level features and high-level color concepts. E.g. with the visual semantics concept "sky" one might assign additional concepts such as "cyan" or "grey" characterizing its color. Also, as the specification of relations between visual entities improves retrieval performance [11], not only do we characterize visual semantics, but also spatial relations linking them. For this, we consider an efficient operational model that allows relational indexing and is adaptable to symbolic image retrieval: conceptual graphs [13]. In the remainder of this paper, we first present the general organization of our model and its representation formalism. We will deal in sections 3, 4 and 5 with the descriptions of the visual semantics, signal and spatial facets. Section 6 will finally specify the query framework.

[1] *Semantics/Signal integration for image retrieval*

2. The SIR Model And Representation Formalism

In state-of-the-art CBIR systems, images cannot be easily or efficiently retrieved due to the lack of a comprehensive image model that captures the structured abstractions, the conveyed signal information and the image semantic richness. To remedy such shortcomings, visual semantics, signal and spatial features are integrated within an image model (figure 1) which consists of a physical and a conceptual image level.

Figure 1. Image model.



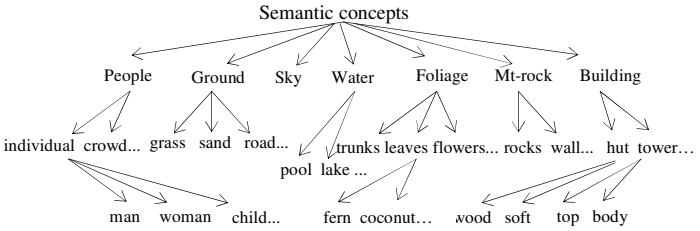
The first layer of the SIR image model (figure 1) is the physical image level representing an image as a matrix of pixels. The second layer of the model is the conceptual layer and is itself a tri-facetted structure (visual semantics, signal and spatial facets). An image is characterized as a set of **image objects** (IOs), abstract structures representing visual entities. Their specification is an attempt to operate image indexing and retrieval operations beyond trivial low-level processes [12] or region-based techniques [2] since IOs convey the visual semantics, signal and spatial information at the conceptual level. In figure 1, the first IO (Io1) is linked to the visual semantics concept *child* and the signal (color) features *skin*, *black* and *red*. It is *inside* (*a part of*) and at the *center* of the visual entity represented by Io2. Visual semantics, signal and spatial facets will be thoroughly described in sections 3, 4 and 5. The conceptual level is supported by a representation formalism which allows to instantiate it within an image retrieval framework: conceptual graphs (CGs). Index and query images are represented by CGs and the matching function evaluating their similarity is based on the CG projection operator [13]. The asset of this formalism is its flexible adaptation to the symbolic approach of image retrieval [10,11].

Formally, a CG is a finite, bipartite, connex and oriented graph. It features two types of nodes: the first one represented by a rectangle is tagged by a concept however the second represented by a circle is tagged by a conceptual relation. For example, the graph Eating ← Action ← Man → Location → Restaurant represents a man eating in a restaurant. Concept and conceptual relation are organized within a lattice structure partially ordered by '≤' which expresses the relation 'is a specialization of'. For example, person ≤ man denotes that the concept *man* is a specialization of the concept *person*, and will therefore appear in the offspring of the latter within the lattice classifying these concepts.

3. Modeling the Image Semantic Content through the Visual Semantics facet

Semantic concepts are learned and then automatically extracted given a visual vocabulary. The construction of the vocabulary is strongly constrained by the application domain [10] and we deal in this paper with corpus of personal photographs. Starting from a set of 7 semantic concepts representing the main visual entities within personal photographs (*people*, *ground*, *sky*, *water*, *foliage*, *mountain/rocks* and *building*), we use WordNet to produce a list of hyponyms linked to these concepts and discard terms which are not relevant for our purpose. We obtain a set of concepts which are specializations of these 7 semantic concepts. We then repeat the process of finding hyponyms for all the specialized concepts. The last step consists in organizing these concepts (72) within a multi-layered lattice of semantic concepts ordered by a specific/generic partial order. In figure 2, the second layer of the lattice consists of concepts which are specific to the major semantic concepts, e.g. *individual* and *crowd* are specific concepts of *people*. The third layer is the basic layer and presents the most specific concepts, e.g. *man*, *woman*, *child* are specific concepts of *individual*. As a matter of fact, Io2 will be represented by the semantic concept *water* and an instance of the visual semantics facet is represented by a set of CGs, each one containing an *Io* type (representing an image object) linked through the conceptual relation *sct* to a semantic concept. E.g., graphs Io1 → sct → Child and Io2 → sct → Water are the representation of the visual semantics facet in figure 1 and can be translated as: the first IO (Io1) is associated with the semantic concept *child* and the second image object (Io2) with the concept *water*.

Figure 2. Lattice of Semantic Concepts.



4. From Low-level Features to the Signal Color Facet

Integrating signal features within a high-level conceptual framework is not straightforward. The first step consists in specifying conceptual signal data which correspond to low-level features. We consider 11 color categories (cyan, white, green, grey, red, blue, yellow, purple, black, skin, orange) characterized in the HVC perceptually uniform space by a union of brightness, tonality and saturation intervals [5]. Each IO is then indexed by a quantified signal index concept (*QSIC*) which features its color distribution by a conjunction of color categories and their corresponding integer pixel percentages. Io1 and Io2 are characterized by QSICs <c:0,s:60,b:25,r:15...> and <c:85,s:0,b:15,r:0...>, which is translated by Io1 having a signal distribution including 60% of skin color, 25% of black and 15% of red; Io2, 85% of cyan and 15% of black.

An instance of the signal color facet is represented by a set of CGs, each one containing an *Io* type linked through the conceptual signal relation q_si to a QSIC. The following CGs are the representation of the signal facet in figure 1: [Io1] → (q_si) → [<c:0,s:60,b:25,r:15...>] [Io2] → (q_si) → [<c:85,s:0,b:15,r:0...>]

5. Modeling Spatial Relations through the Spatial Facet

In order to model spatial data, we first consider a subset of the topological relations explicited in the RCC-8 theory [3]; 3 relations which are exhaustive and relevant for image querying are chosen. Considering 2 IOs (Io1 and Io2), these relations are (**P**,Io1,Io2): 'Io1 is a part of Io2', (**T**,so1,so2): 'Io1 touches Io2 (externally connected or overlaps)' and (**D**,so1,so2): 'Io1 is disconnected with Io2'. Let us note that they are mutually exclusive and present the important property that each pair of IOs is linked by one and only one topological relation. Directional relations **Right**(R), **Left**(L), **Above**(A), **Below**(B) and the global relation **Center**(C) are invariant to basic geometrical transformations (translation, scaling).

An image object is therefore characterized by its centre of gravity *io_g* as well as two pixel sets: its interior, noted *io_i* and its boundary, noted *io_b*. To deal with the computation of topological relations [4], two image objects Io1 and Io2 are characterized by intersections of their interior and boundary sets: io1_i ∩ io2_i, io1_i ∩ io2_b, io1_b ∩ io2_i and io1_b ∩ io2_b. Each topological relation is mapped to the results of these intersections, e.g. (DC, so1, so2) iff. io1_i ∩ io2_i =∅, io1_i ∩ io2_b =∅, io1_b ∩ io2_i =∅ and io1_b ∩ io2_b =∅. The interest of this computation method relies on the association of topological relations to the previous set of necessary and sufficient conditions involving attributes of spatial objects (i.e interior and boundary). The computation of directional relations between Io1 and Io2 relies on the relative position of their centers of gravity.

An instance of the spatial facet is represented by a set of CGs, each one containing 2 *Io* types linked through the previously defined spatial relations. The following CGs are the representation of the spatial facet in fig. 1 and can be translated as Io1 overlaps and is at the center of Io2: [Io1] → (T) → [Io2] [Io1] → (C) → [Io2]

6. The Query Module

Our architecture is based on a unified fully textual framework allowing a user to query over the visual semantics, signal and spatial facets. This obviously enhances user interaction since contrarily to state-of-the-art systems, the user becomes in 'charge' of the query process by making his needs explicit to the system

through full textual querying (cf. [1] for further details). We propose an expressive and computationally efficient language consisting of boolean and quantified queries:
- A user is able to associate visual semantics concepts and their related spatial relations with a boolean conjunction of color categories in a query such as *Find images with a child with (i.e. wearing) red **and** black inside cyan water*.
- 'At Least' queries (*Find images with a child inside pool water (At Least 50% of cyan)*) and 'At Most' queries (*Find images with a child inside lake water (At Most 25% of cyan)*) associate visual semantics concepts and their related spatial relations with a set of color categories and a percentage of pixels belonging to each one of these categories. We specify also literally quantified queries (*Mostly*, *Few*), related to numeral quantifications 'At least 50%' and 'At most 10%' which are easier to handle by a user not interested in querying precisely with percentages.
Each image (respectively user query) is represented by a global CG resulting from the aggregation of CGs over the visual semantics, signal and spatial facets called document index graph (respectively query graph). The evaluation of similarity between an image and a query is achieved through a correspondence function: the CG projection operator. This operator allows to identify within a graph g_1 sub-graphs with the same structure as a given graph g_2, with nodes being possibly restricted, i.e. their types are specialization of g_2 node types. If it exists a projection of a query CG Q within a document CG D then the document indexed by D is relevant for the query Q. As far as implementation is concerned, optimizations related to the organization of index data structures have been developed allowing to process this operator in polynomial time within a given application domain [11].

5. CONCLUSION

We have specified a multi-facetted architecture for image retrieval instantiated by an operational model based on the CG formalism. The latter allows to define an image representation model and a matching function to compare index and query image graphs. We have specified image objects, abstract structures representing visual entities within an image in order to operate image indexing and retrieval operations at a higher level of abstraction than state-of-the-art frameworks. We have described the visual semantics, signal and spatial facets that characterize the conceptual information conveyed by image objects and have finally proposed a unified and rich framework for querying over visual semantics, signal and spatial features.

REFERENCES

[1] Belkhatir, M. & Mulhem, P. & Chiaramella, Y. "Integrating Perceptual Signal Features within a Multi-facetted Conceptual Model for Automatic Image Retrieval". ECIR, pp. 267-282, 2004
[2] Carson, C. & al. "Blobworld: A System for Region-Based Image Indexing and Retrieval". VISUAL, 509-516, 1999
[3] Cohn, A. & al. "Representing and Reasoning with qualitative spatial relations about regions". Chap. 4, 97-134, 1997
[4] Egenhofer, M.J. "A formal definition of binary topological relationships". Int. Conf. on foundations of data organization and algorithms, 457-472, 1989
[5] Gong, Y. & al. "Image Indexing and Retrieval Based on Color Histograms". Multimedia Tools and App., 133-156, 1996
[6] La Cascia & al. "Combining Textual and Visual Cues for Content-Based IR on the World Wide Web". IEEE Workshop on Content-Based Access of Image and Video Libraries, 24-28, 1998
[7] Lim, J.H. "Explicit query formulation with visual keywords". ACM MM, 407-412, 2000
[8] Lim, J.H. & al. "Home Photo Content Modeling for Personalized Event-Based Retrieval". IEEE MM 10(4), 28-37, 2003
[9] Lu, Y. & al. "A unified framework for semantics and feature based RF in image retrieval systems". ACM MM, 31-37, 2000
[10] Mechkour, M. "EMIR^2: An Extended Model for Image Representation and Retrieval". DEXA, 395-404, 1995
[11] Ounis, I. & Pasca, M. "RELIEF: Combining expressiveness and rapidity into a single system". SIGIR, 266-274, 1998
[12] Smeulders, A. & al. "Content-based image retrieval at the end of the early years". IEEE PAMI 22(12), 1349-1380, 2000
[13] Sowa, J.F. "Conceptual structures: information processing in mind and machine". Addison-Wesley, 1984
[14] Zhou, X. & Huang, T.S. "Unifying Keywords and Visual Contents in Image Retrieval". IEEE MM 9(2), 23-33, 2002